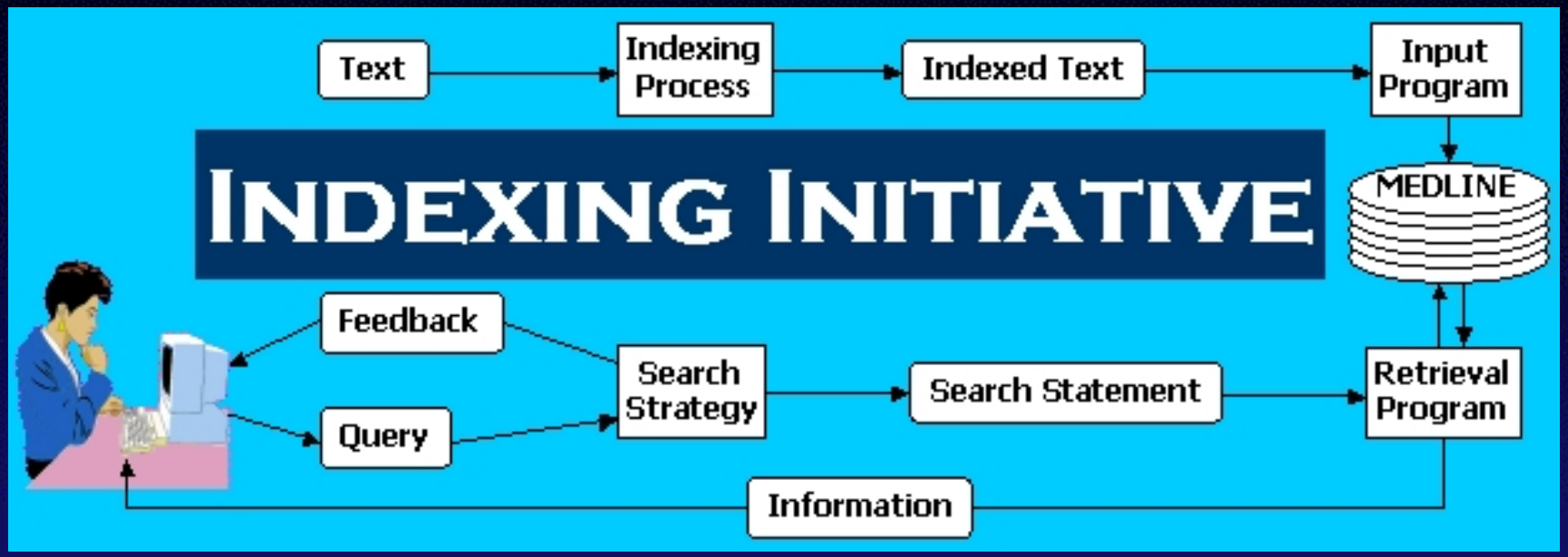
Automated and Semi-automated Indexing

Alexa T. McCray
Alan R. Aronson

Board of Regents

September 11, 2002

U.S. NATIONAL LIBRARY OF MEDICINE



U.S. NATIONAL LIBRARY OF MEDICINE

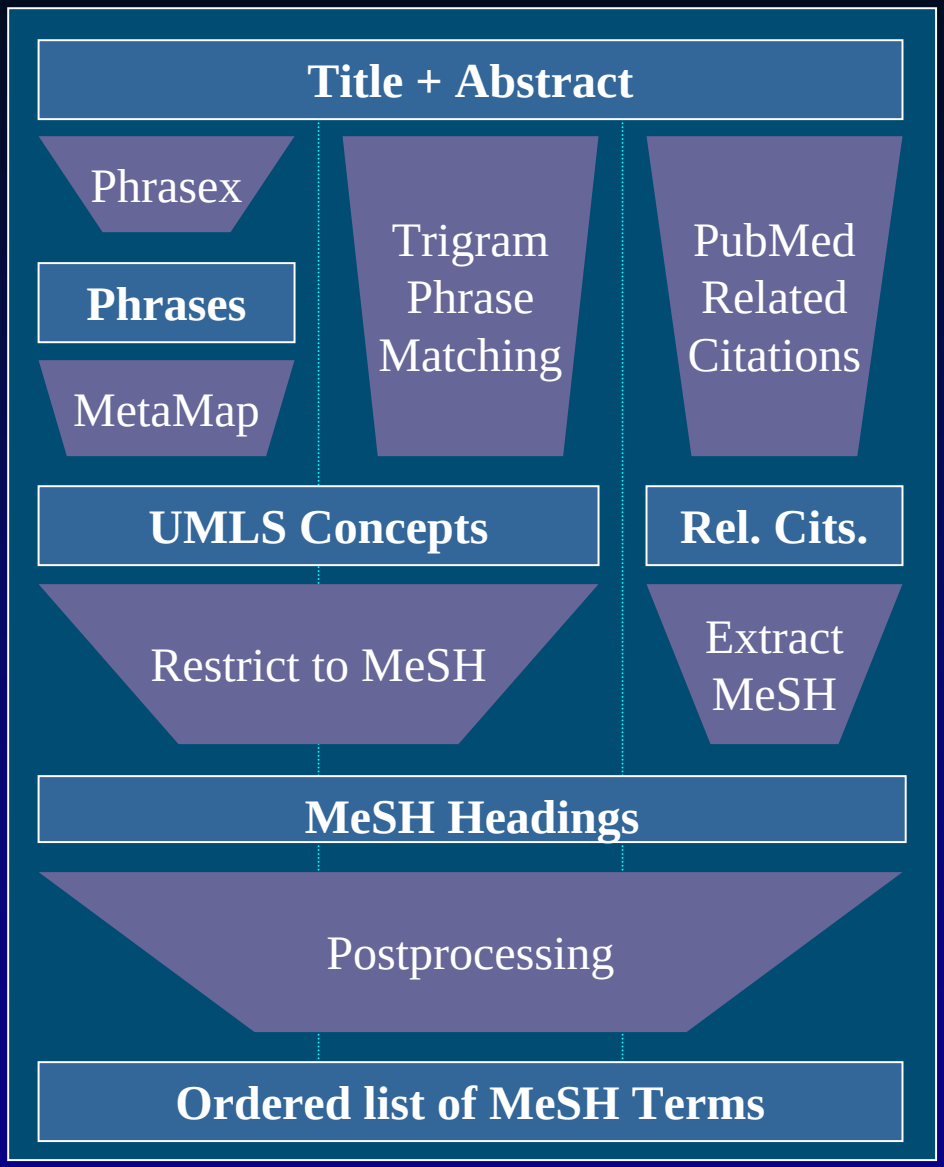
Indexing Initiative (II) Project Goals

- Investigate automated and semi-automated indexing methodologies
- Develop methods that result in acceptable retrieval performance
  - Concept-based algorithms
  - Extensive use of UMLS resources

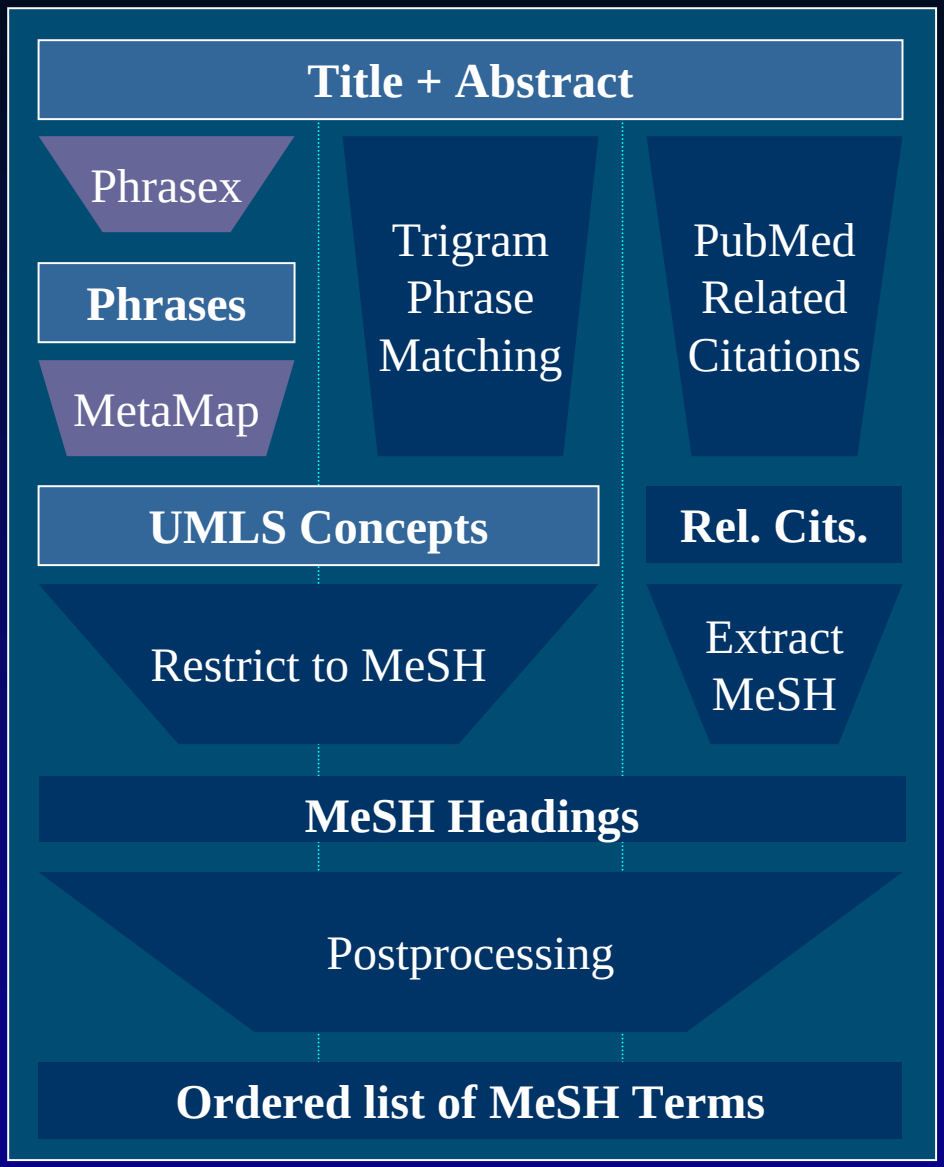
II Project Phases

1. Initially, an independent collection of projects addressing
   - Indexing methods
   - Evaluation
   - Policy
2. Development of a prototype indexing system for testing indexing methods
3. Deployment of the Medical Text Indexer (MTI) system to NLM indexing environments

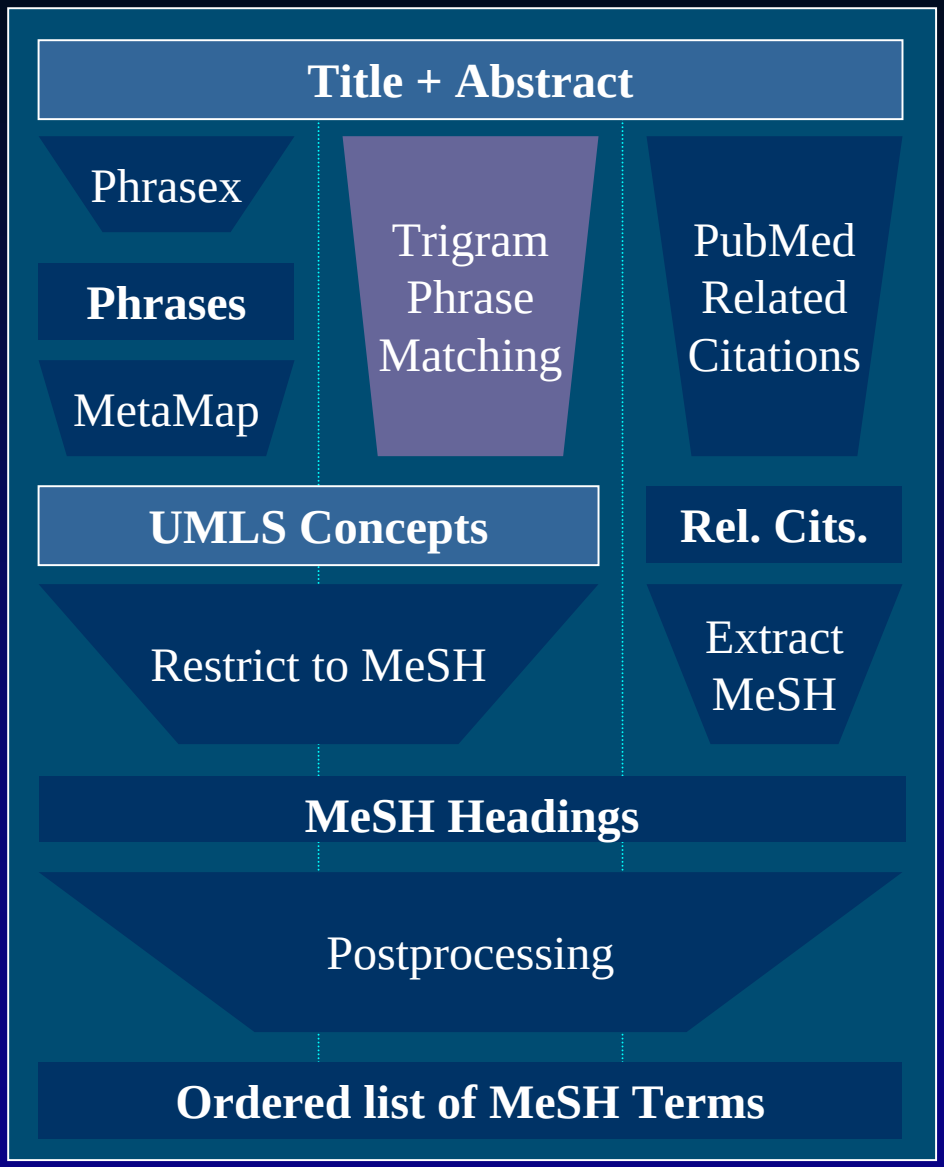
The Medical Text Indexer (MTI)



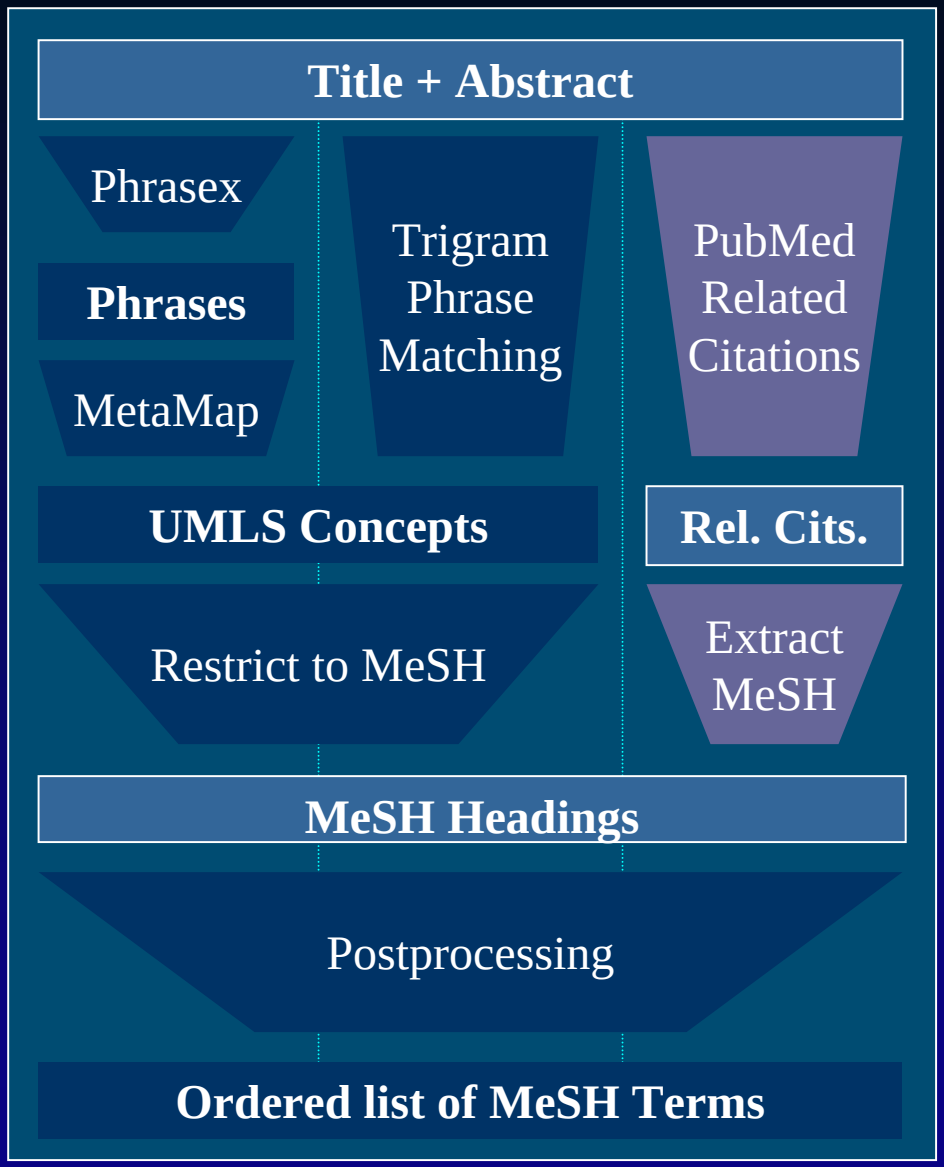
MetaMap Indexing



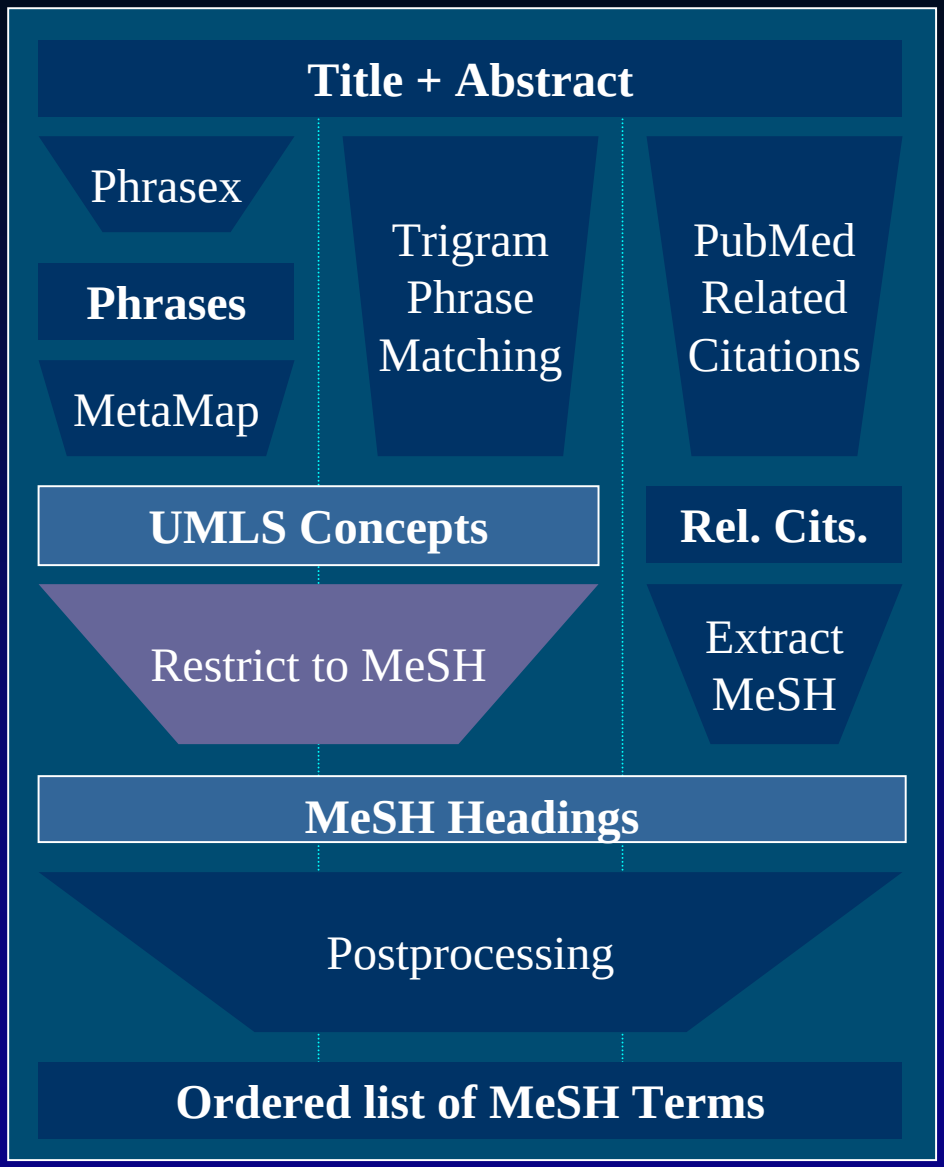
Trigram Phrase Matching



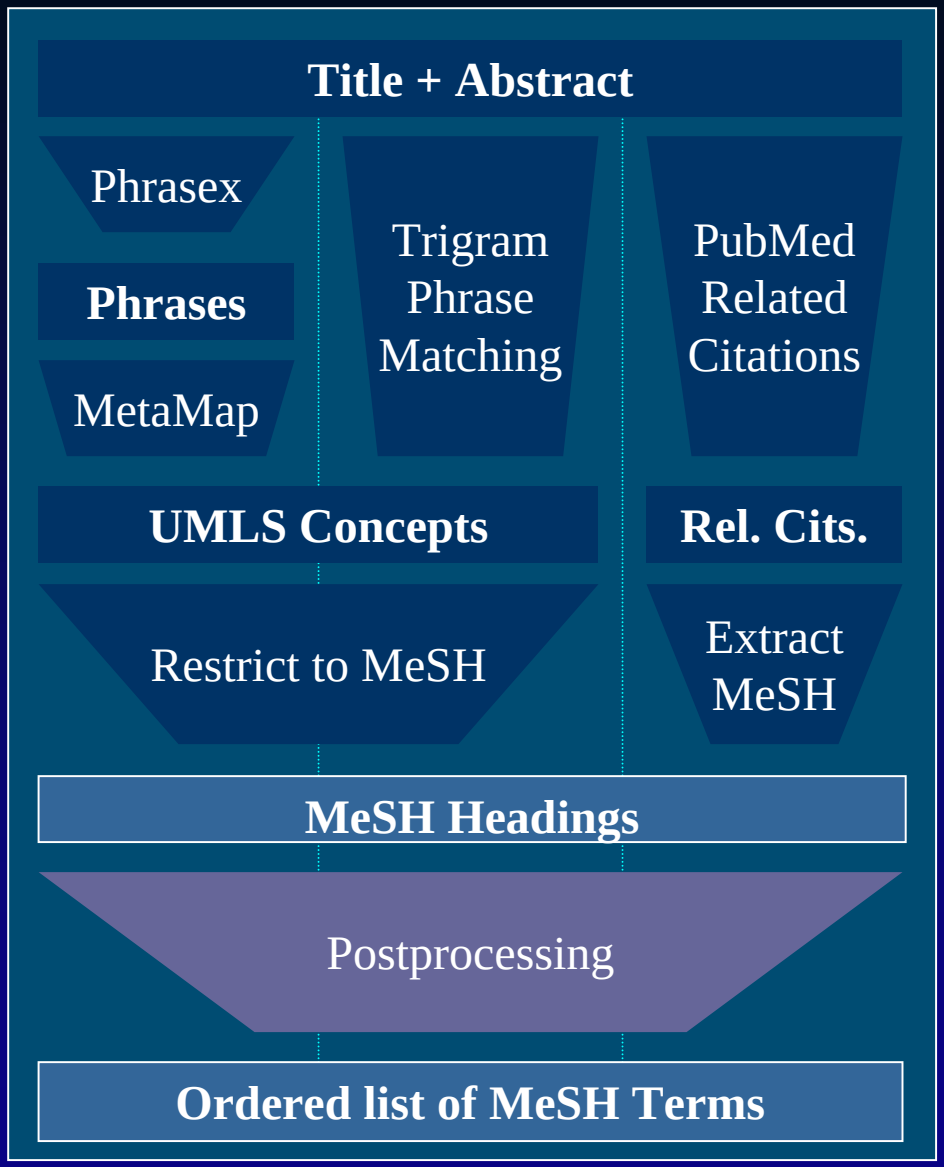
PubMed Related Citations



Restrict to MeSH



Postprocessing



Phrase-based Indexing Methods

- MetaMap Indexing
  - Perform MetaMap processing on input text
    - Parse text into phrases
    - Generate variants
    - Retrieve Metathesaurus candidates
    - Retrieve and evaluate Metathesaurus candidates
    - Construct final mapping
  - Rank all concepts discovered
- Trigram phrase matching
  - Form phrases based on character trigrams
  - Match against Metathesaurus

MetaMap Example

- Text: *"The local anesthetic bupivacaine is cardiotoxic …"*
- Phrases: "The local anesthetic bupivacaine", "is", "cardiotoxic", …
- Variants: *anesthetics*, *anaesthetic*, *anesthesia*, …
- Candidates: 'Bupivacaine', 'Local anaesthetic', 'Local anaesthetic, NOS', …
- Mappings
  - 'Bupivacaine' and
  - 'Local anaesthetic' or 'Local anaesthetic, NOS'

PubMed Related Citations Indexing

- Find the closest neighbors (related citations) to the input text
- Extract the MeSH headings from the neighbors
- Example
  - Text: “Bupivacaine inhibition of L-type calcium current in ventricular cardiomyocytes of hamster. …”
  - Extracted MeSH:
    - ‘Calcium Channels’
    - ‘Calcium Channel Blockers’

Restrict to MeSH

- Find the semantically closest MeSH headings using UMLS relationships:
  - Synonyms
  - Associated expressions
  - Hierarchical relationships (child, parent)
  - Other relationships
- ‘Acute adenoviral follicular conjunctivitis’ restricts to
  - ‘Adenoviridae Infections’ and
  - ‘Conjunctivitis, Viral’

Postprocessing (1 of 2)

- Clustering of results from basic methods
- Indexing rules and lookup lists
  - ‘Eclampsia’ -> ‘Female’ and ‘Pregnancy’
  - G05 treecode -> ‘genetics’
  - “pediatric(s)” -> ‘Child’
  - ‘Hamsters’ -> ‘Animal’
- Exclusions (e.g., ‘TEST’, ‘Disease’)
- Further promotion of title headings and chemicals

Postprocessing (2 of 2)

- UMLS/MeSH heuristics
  - Remove MM heading with unrelated semantic type
  - Remove RC heading if no more general MM heading
  - Remove a chemical MM heading when no other terms are chemical in nature

MM – MetaMap recommendation
RC – Related Citations recommendation

A MEDLINE Citation

TI - Bupivacaine inhibition of L-type calcium current in ventricular cardiomyocytes of hamster.

AB - BACKGROUND: The local anesthetic bupivacaine is cardiotoxic when accidentally injected into the circulation. Such cardiotoxicity might involve an inhibition of cardiac L-type Ca2+ current (ICa,L). This study was designed to define the mechanism of bupivacaine inhibition of ICa,L. … CONCLUSIONS: The inhibition of ICa,L appears, in part, to result from bupivacaine predisposing L-type Ca channels to the inactivated state. Data from washout suggest that there may be two mechanisms of inhibition at work. Bupivacaine may bind with low affinity to the Ca channel and also affect an unidentified metabolic component that modulates Ca channel function.

Assigned MeSH and Suggested MTI Terms

- Assigned MeSH (10)

  ***Anesthetics, Local**

  Animal

  ***Bupivacaine**

  ***Calcium Channels**

  Calcium Channels, L-Type

  Dose-Response Relationship, Drug

  Hamsters

  *Heart

  Male

  Support, Non-U.S. Gov't

- Suggested MTI Terms (11)

  1. Calcium
  2. Heart Ventricle
  3. **Bupivacaine**
  4. **Calcium Channels**
  5. Calcium Channel Blockers
  6. **Calcium Channels, L-Type**
  7. Cells
  8. Calcium Channels, T-Type
  9. **Anesthetics, Local**

  Hamsters

  Animal

MTI Deployment: Fully Automated Indexing

- MTI indexing of collections which will not be manually indexed
- Initial collections available from the NLM Gateway
  - AIDS/HIV meeting abstracts
  - Health services research meeting abstracts
  - Space life sciences meeting abstracts
- System in final testing before deployment

Evaluation of Fully Automated Indexing

- Retrieval experiments together with
- Continued system development to improve accuracy
  - Basic MTI components
  - Word Sense Disambiguation (WSD) research

MTI Deployment: Semi-automated Indexing

- MTI recommendations presented to indexers within the Data Creation and Maintenance System (DCMS)
- Live experiment with volunteer indexers
- Using feedback from experiment to improve the quality of MTI recommendations

MTI Indexing Experiment

- Ten volunteers each indexed a journal issue using MTI recommendations
- Questionnaires for each article indexed plus summary questionnaire
- Analysis
  - Average of 8 useful terms per article (3 main)
  - Precision = .29, Recall = .55
  - Adequate coverage? 37% yes, 53% partial, 10% no

Experiment Feedback

- Make suggested terms hot links to the MeSH browser
- Gray out selected terms
- Show entry term, not heading, if found
- Provide interactive access to MTI

Evaluation and Status of Semi-Automated Indexing

- Evaluation
  - Comparison of final indexing with MTI suggestions
  - Further feedback after implementation of indexers recommendations
- Deployed for all indexers August 29, 2002

Indexing Initiative Contributors

- LHNCBC
  - Alan R. Aronson
  - Olivier Bodenreider
  - Clifford W. Gay
  - William T. Hole
  - Susanne M. Humphrey
  - James G. Mork
  - Alexa T. McCray
  - Thomas C. Rindflesch
  - Will J. Rogers
  - Sonya E. Shooshan
- NCBI
  - Won Kim
  - W. John Wilbur
- OCCS
  - John Butler
  - John M. Rozier
- LO
  - Ione Auston
  - Nadine Benton
  - Andrea Demsey
  - Lou S. Knecht
  - James R. Marcetich
  - Stuart J. Nelson
  - Marina P. Rappoport
  - Jane L. Rosov
  - Catherine R. Selden
  - Sara J. Tybaert
  - Joe D. Thomas
  - Carolyn B. Tilley
  - Janice M. Ward
- SIS
  - H. Florence Chang
  - Tamas E. Doszkocs
  - George (Mike) F. Hazard

A MEDLINE Citation (Full version)

TI - Bupivacaine inhibition of L-type calcium current in ventricular cardiomyocytes of hamster.

AB - BACKGROUND: The local anesthetic bupivacaine is cardiotoxic when accidentally injected into the circulation. Such cardiotoxicity might involve an inhibition of cardiac L-type Ca2+ current (ICa,L). This study was designed to define the mechanism of bupivacaine inhibition of ICa,L. METHODS: Cardiomyocytes were enzymatically dispersed from hamster ventricles. Certain voltage- and time-dependencies of ICa,L were recorded using the whole-cell patch clamp method in the presence and absence of different concentrations of bupivacaine. RESULTS: Bupivacaine, in a concentration-dependent manner (10-300 microM), tonically inhibited the peak amplitude of ICa,L. The inhibition was characterized by an increase in the time of recovery from inactivation and a negative-voltage shift of the steady-state inactivation curve. The inhibition was shown to be voltage-dependent, and the peak amplitude of ICa,L could not be restored to control levels by a wash from bupivacaine. CONCLUSIONS: The inhibition of ICa,L appears, in part, to result from bupivacaine predisposing L-type Ca channels to the inactivated state. Data from washout suggest that there may be two mechanisms of inhibition at work. Bupivacaine may bind with low affinity to the Ca channel and also affect an unidentified metabolic component that modulates Ca channel function.